DOCUMENT RESUME

ED 151 410 TM 006 933

AUTHOR Davis, Barbara Gross
TITLE Evaluating Innovative Programs: The Berkeley Perspective.
PUB DATE Apr 76
NOTE 27p.; Paper presented at the Annual Meeting of the American Educational Research Association (60th, San Francisco, California, April 19-23, 1976)

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS *Affective Objectives; *Basic Skills; Cultural Pluralism; Decision Making; *Educational Alternatives; Educationally Disadvantaged; Elementary Secondary Education; Ethnic Groups; *Evaluation Methods; *Evaluation Needs; Evaluators; Experimental Programs; Experimental Schools; Formative Evaluation; *Program Evaluation; Racial Integration; Research Problems; Summative Evaluation
IDENTIFIERS California (Berkeley)

ABSTRACT

The Berkeley Experimental Schools Project (BESP) consolidated the existing elementary and secondary educational alternatives with promising new programs in the Berkeley Unified School District of California. The program goals involve the elimination of racism, acquisition of basic skills, affirmation of cultural pluralism, and shared decision making. The internal evaluation team provided information to the program director for decision making and dissemination to parents, faculty, staff, and students. The external evaluation team was responsible to the National Institute of Education (NIE), and assessed the impact of BESP on the district as a whole. This paper highlights what has been learned about internal evaluation functioning as a result of BESP. Two general aspects are discussed: (1) the administrative or organizational concerns of an evaluation unit--staffing requirements and location of the evaluation component in relation to the district structure, and (2) evaluation strategies for assessing educational alternatives--selecting the appropriate methodology and determining the successful outcomes. (Author/MV)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

ED151419

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Barbara Gross Davis

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

# EVALUATING INNOVATIVE PROGRAMS:

# THE BERKELEY PERSPECTIVE

By

Barbara Gross Davis

Office of Research and Evaluation

Berkeley Unified School District
1414 Walnut Street
Berkeley, California 94703

Paper presented at the annual meeting of the

American Educational Research Association,

San Francisco, California, April, 1976

TM006 933

Description of Berkeley Experimental Schools Project

Alternative education programs began in Berkeley in response to students and parents asking for choices in teaching styles, philosophy, curriculum and goals. The first alternatives in the late sixties were efforts to create smaller schools for students as options to the large high school campus. In 1971 when grant money was made available to Berkeley Unified School District and other school districts for five years to develop educational programs to affect major changes, the existing alternatives in Berkeley as well as promising new programs were consolidated into the Berkeley Experimental Schools Project (BESP).

The original goals of BESP were:

1. To provide a system which could move toward the elimination of racism in the schools and the larger community.

2. To provide a system which would facilitate the acquisition of basic skills with special focus on those who are educationally disadvantaged or who are members of ethnic minority groups.

3. To provide significant changes in the administration and organization of the system so that power of decision making could become a shared activity.

4. To provide program options that would foster the cultural pluralism extant in the school community and affirm the district's value of it.

As a result of the size of the grant and the diversity of educational views and philosophy in Berkeley, BESP originally encompassed 23 different educational programs from pre-school to high school. The programs were of three general types: off-site, which were physically separate schools; schools within schools which shared the regular district facilities but had their own administrative structure and school identity, and alternative curriculum offerings which were schools providing options for students in how they learn. It is difficult to identify common characteristics among the 23 programs. They have ranged from supplementary add-on curricula to separate and unique schools. They have covered grades K-12 and served racially and economically diverse populations.

In the final two years of funding the number of programs has been reduced to 14. Three of the original programs, Black House, Casa de la Raza and UN West, were suspended by order of the Office of Civil Rights because they involved students primarily of only one racial or ethnic group. Two other programs were phased into the regular school curriculum (disbanded). Two programs were merged; and four programs were consolidated under one administrative structure.

The 14 remaining programs cover all three general alternative school types and all K-12 grades. Approximately 3500 students each year out of a total school population of about 12,000 have elected to be involved in BESP.

## BESP Level I Evaluation: A Brief History

The evaluation of BESP has been both internal (called Level I) and external (called Level II). The Level I team provides information about the project to the Director of BESP for program decision making and dissemination to parents, faculty, staff, and students. The Level II team is responsible to National Institute of Education (NIE) and assesses the impact of BESP on the district as a whole. Although in the original conception Level I was to provide formative evaluation information and Level II was to provide summative data, in practice this distinction has become blurred. Level I has had a shifting focus and the last two years has operated more in the summative than the formative mode.

The BESP Level I team in Berkeley has undergone three major turnovers in staffing, focus and functioning in four years. The first Level I was a team of eight responsible to a Director of BESP Evaluation. They were independent of the district Office of Research and Evaluation. In the first years of the project Level I was engaged largely in simply helping the many programs get on their feet and underway; evaluation provided assistance to facilitate program implementation. There was no specific focus for Level I efforts and no attempt to collect systematic data or

document evaluation activities.

The second Level I team consisted of 14 people under a new Director of BESP Evaluation. This group was involved in exploring new methods and modes of evaluation. Their activities were unique at each site, and only a limited amount of common data was collected across sites (such as socio-economic status of participants, racial distribution of BESP participants, and an attempt at rating the "alternativeness" of each school).

Evaluation in the early years of BESP was difficult for several reasons: first, the state of the art of evaluation in general was not well developed. An ERIC search in 1972 revealed only 27 references for evaluating educational innovations. Many of those references were general books or specific articles in particular content areas (higher education, computer assisted instruction, etc.). These references seemed to have little relevance for the Berkeley situation.

Besides the lack of concrete techniques and procedures for formative evaluation of educational alternatives, there also existed the widespread attitude among the BESP programs that evaluation could not effectively assess their efforts. The common complaint was, "What we are doing can't be measured by questionnaires, interviews or even observations. We've created a climate, a feeling, an intangible sense of community that makes our program work and evaluation can't capture that essence." Several sites did not participate in the district-wide achievement testing in the early years, and no baseline data were collected.

Finally, the formulation of some of the BESP goals into measurable objectives proved impossible. Terms like "institutional racism" or "cultural pluralism" have not yet been adequately defined and consequently only tangentially measured. Thus because of the lack of concrete formative evaluation methodology, the resistance of the sites to evaluation (which they equated with tests and measurement) and the difficulties in quantifying BESP goals in measurable terms, Level I failed to adequately assess the process or impact of BESP.

The third Level I team was constituted for the 1974-75 school year and consisted of four staff members under the auspices of the district Director of Research and Evaluation. For the first time the evaluation component was placed within the regular district structure. It was by this time the next to the last year of the project. With only two remaining years the Director of BESP, and district administration, were concerned about collecting some concrete, detailed information on the impact of BESP. Had it made a difference in basic skill achievement? Had it had any affective impact on students? Had power-sharing become a reality? How did teachers and parents feel about the educational alternatives? After much negotiation another evaluation plan was accepted by NIE and implemented by the new Level I. This final plan relied on traditional evaluation procedures (treatment versus comparison group; survey design) and established educational outcomes (basic skill achievement; student self-concept, attitude toward school, perceptions of

ethnic relations and feelings about ESP; staff attitudes and expectancies; parent attitude and expectancies). Because of the small size of the staff and the summative focus of the evaluation design, only limited assistance has been provided to the 14 individual ESP programs. The unique aspects of each program are not being adequately evaluated; rather all programs are assessed on the same general criteria.

In Berkeley, the original NIE intention of having a formative evaluation unit never really materialized. Some formative evaluation was conducted in the early years but in a rather haphazard unsystematic fashion. Summative evaluation predominated in the later years but tended to ignore the unique aspects of each program. The purpose of the remainder of this paper is not to examine the reasons for the Level I changes but rather to identify some of the key issues involved in evaluating alternative educational programs as they emerged in the Berkeley experience. The remarks that follow do not necessarily represent the opinions of others associated with BESP Level I or the district as a whole, but reflect my personal perspective, based on an association with the project only in its last two years.

## Issues in the Evaluation of Educational Alternatives

The BESP experience embodied five years of complicated educational issues, inflamed emotions, and complex sociological events. It is beyond the scope of this presentation to deal in detail with the entire BESP evaluation experience (the interested

reader is referred to Jones (in progress) for a more in-depth examination of BESP). Rather my intent is to highlight something of what we learned about internal evaluation functioning as a result of BESP.

The focus will be on two general aspects: the administrative or organizational concerns of an evaluation unit (staffing requirements and location of the evaluation component in relation to district structure) and evaluation strategies for assessing educational alternatives (selecting appropriate methodolology and determining the "success" of an alternative). These topics, of course, do not exhaust the range of issues involved in evaluating educational alternatives. However, these areas were selected because of their importance in actually setting up and implementing an evaluation component.

Staffing requirements of an internal evaluation component

The issue of who should evaluate educational alternatives has at least three facets. The first is a question raised by Worthen (1974): Should the evaluator be a content specialist in the program under evaluation (someone with expertise in open education or individualized learning, etc.) or an evaluation specialist (someone with evaluation training or experience)? Worthen came to the conclusion that on balance the evaluation specialist is preferable. He feels the content of most educational areas is relatively easy to grasp and the evaluator is apt to be spared the intrusion of personal biases affecting the

content specialist. The early Level I teams had only limited experience in educational evaluation which may have contributed to some of the difficulties they encountered in assessing the alternative programs. The Berkeley experience has underscored the importance of having a staff with the necessary evaluation skills.

The second factor in the who-should-evaluate-educational-alternatives issue is what has been called by Patton (1975) the "distance vs. closeness" conflict. Evaluation has an aura of objectivity, based, in part, upon its ties to the scientific method. An evaluator is usually seen as an objective impartial judge of the effectiveness of a program who keeps a psychological distance from the project under evaluation. Recently several researchers (Patton, 1975; Hein, 1975) have argued against distance and detachment, particularly for innovative programs. They believe that the evaluator should strive for psychological closeness to the program through physical proximity and through the development of an atmosphere of intimacy and confidentiality with the staff. This view is rejected by those who feel social bonds may limit the objectivity of the evaluation, destroy the external credibility of the evaluation and perhaps invalidate evaluative judgments altogether (Scriven, 1974a).

The issue for the evaluator is how involved to become with the program and staff. In this context "close" refers to more than the usual evaluation procedures of observation, interviews, and knowledge about the program. "Close" involves a psychologi-

cal identification with and commitment to the project under evaluation. There are several perceived advantages to becoming closely allied with the program:

1. Resistance to evaluation may be minimized because the evaluator becomes identified with the staff and is not viewed as a potentially hostile outsider.

2. Intense involvement with the program may provide a rich source of descriptive or supplemental data not otherwise obtainable.

3. Program participants may recognize as legitimate only evaluations done by persons close to the program.

On the other hand, the potential drawbacks are:

1. Someone close to the program is likely to be less independent and objective than someone at a greater remove because of, perhaps, social pressures, feelings of loyalty or identification with program.

2. An evaluator closely associated with the program under evaluation may become enmeshed in its internal politics and forfeit his/her effectiveness with the staff as a whole.

3. An evaluator closely associated with the program may have less credibility with the policy makers or the public than someone with more distance.

One obvious solution to the distance-closeness conflict is to have two evaluators for every program; one to be "close" and one to remain distant. Presumably the respective biases and insights of each type of evaluator would combine to form a total picture of program functioning. This, in fact, may have been the rationale behind the Level I and Level II dichotomy as conceived by NIE. But in Berkeley, the limited resources (four people for 17 programs), the type of data requested by decision makers, the concerns about the credibility of results, and the elimination of two previous teams supposedly engaged in being close to the program all militated toward a more psychologically distant posture for the third Level I team.

The distance-closeness conflict may not be an either/or situation. Obviously the personality of the evaluator, the type of program, the feelings of the staff and other variables will influence how close to a program an evaluator becomes. The crucial issue is for the evaluator to realize when he/she has lost objectivity or credibility. Scriven (1974$_a$) has recommended procedures to identify deterioration of objectivity of evaluators and provide support or replacement when indicated. This monitoring can be most easily accomplished by providing a system for continuous evaluation of the evaluator's work.

The third area of debate in who should evaluate alternatives involves the use of interdisciplinary teams. All three Berkeley Level I teams have been essentially interdisciplinary. Team members' backgrounds included teaching, educational administration,

political science, sociology, psychology, and evaluation. In addition some team members have seen themselves as psychologically close to BESP and committed to alternative education while others have maintained a neutral posture about alternative education. This diversity might have produced a balanced evaluation team but it has in fact produced a fragmented unit. Petrie (1976) has documented some of the features of a successful interdisciplinary team, notably: a clearly recognizable central focus for the effort, disciplinary competence by each member of the team, personality characteristics such as adventurous spirit, and broad interests, strong administrative support, realistic expectations of what can be accomplished, etc. Few of these features have been fully realized in the BESP Level I teams.

In setting up evaluation teams, two factors are crucial. Staff must have demonstrated proficiency in evaluation methods and techniques. Specialists in certain areas can be used as consultants or staff members, but the overall responsibility for team functioning should rest with someone competent in evaluation. Second, the professional values of the team (preference for case study or experimental design; product or process orientations, etc.) should be made explicit and critically examined (Ball & Anderson, 1975). Professional values affect the choice of evaluation strategies, methods and procedures. Evaluation teams with widely divergent professional values may work neither efficiently nor as effectively as a more homogeneous team bringing in divergent viewpoints through the use of outside consultants.

Location of the Project Evaluation Unit

Berkeley Level I may be unique in that it started out as part of the organizational structure of BESP, as a separate entity from the small but functioning Office of Research and Evaluation, and changed in midstream. For the past two years BESP Evaluation has been part of the district Office of Research and Evaluation. In the early years of BESP, Level I and the Office of Research and Evaluation interacted infrequently and maintained autonomy in evaluation activities. The switch in recent years to being part of the district Office of Research and Evaluation has had both advantages and disadvantages for BESP evaluation. At the beginning there was confusion over the chain of command. Was the unit responsible to the Director of Research and Evaluation or to the Director of BESP? Fortunately because of the informal nature of the office and the relationship between the two directors, few conflicts arose. Nevertheless, there were times when Level I didn't know whose expectations to meet.

Another difficulty was the increased demands placed on Level I. As part of the Office of Research and Evaluation, some BESP evaluation personnel have found themselves assisting in meeting district (non-BESP) evaluation needs. These additional responsibilities are viewed with mixed feelings. On the one hand, they indicate that evaluators are considered useful by the district and capable of providing relevant information. On the other hand, there is sometimes an overwhelming sense of frustration for the

evaluator in being unable to adequately complete the district-wide activities and provide satisfactory services to the BESP schools.

Finally, because BESP is now located within the central office, it may have become identified in the eyes of the school community, with the district Office of Research and Evaluation. People's feelings (good or bad) about that office may have been transferred to BESP Evaluation simply by process of association.

There have been equally distinct advantages to being a part of the central office. BESP Evaluation has been able to have a direct impact on some district evaluation functioning; the reorganization of achievement data files in formats more compatible with program evaluation needs has been one major accomplishment. As part of the Office of Research and Evaluation, BESP Level I has had fewer problems securing suitable comparison groups than might have occurred when it operated separately. And most important, communication with those involved in other evaluation projects has enhanced each group's functioning. This communication might have occurred regardless of where Level I was housed, but sharing common space has facilitated sharing professional concerns.

On balance, a strong central district Office of Research and Evaluation, that can meet the needs of a wide variety of ongoing educational projects, is preferable to separate independent evaluation units. A single central evaluation component that expands appropriately each time a new project is added to

the district offers several advantages:

1. It can reduce redundancy in disparate evaluation activities.

2. It can eliminate bias in evaluation because evaluators all functioning as central staff can be rotated among different educational programs as appropriate (some evaluators might be "close" evaluators for one project and "distant" evaluators for other projects, etc.)

3. It can provide a built-in supportive environment for evaluators, a place for sharing mutual concerns.

"Appropriate" Evaluation Methodologies

Once the evaluation structure has been determined, the next important step in the evaluation process is the selection of an appropriate strategy or methodology. One of the goals of NIE in funding the evaluation components was to encourage the development of new methodologies, new evaluation techniques, and new instruments. The ambition was worthy; the realization of it more difficult. For one thing, the development of new instruments requires a great deal of time and resources. The hours needed to engage in this activity are usually not available to evaluators since the ongoing evaluation activities themselves require substantial time commitments. Second, the methods and boundaries of evaluation procedures are so imprecisely delineated that it is unclear what constitutes a "new" technique. Is,

for example, the utilization of a social psychological perspective for examining the effectiveness of the school structure considered a "new" methodology (Schmuck, 1975)? Is the application of the methods of criminology to deduce causality in the evaluation of programs a "new" procedure (Scriven, 1974b)? Or ought these techniques to be considered part of any evaluator's repertoire? Evaluation encompasses a wide variety of methods and procedures; there is no single "right" way to evaluate a program. Some methods are preferable to others (Bernstein and Freeman, 1975) but, by in large, choice of instrument, choice of design, choice of procedure, etc. are dictated by exigencies of time, resources, decision maker's interests and evaluator's knowledge and expertise. Finally, there is little documentation on the effectiveness of "old existing" procedures. Successful evaluations need to be identified, successful meaning, not necessarily evaluations that indicate program success but evaluations that are themselves considered successful (Stufflebeam, 1973). These "successful" evaluations need to be examined so that the particular techniques and procedures that make them successful can be identified. One hopes some consistency can be found among the wide variety of evaluations revealing some methods"more effective" than others. The fact is, the evaluation field lacks a suitable well-defined set of "guaranteed" methods and procedures for evaluating educational programs.

Despite the problems cited above, in the early years of BESP Level I there was an attempt to develop/use innovative

methods. The results were largely disappointing. Considerable time was wasted in an attempt to create new techniques that ended up not providing useful information. In the meantime the collection of critical baseline data on BESP participants was neglected and as a result there has been only limited data collected consistently from year one of the project.

As was indicated earlier, the evaluation plan for Level I the past two years is a rather traditional evaluation design using existing procedures. It may be that in Berkeley, the logical process was reversed. Level I started out trying to be innovative and ended up rather traditional. The better route would have been to begin with obtaining some baseline data with the usual measures and procedures and then gradually develop "new" methods as the project progresses.

Whether "innovative" methods or "existing" procedures are used, the importance of specifying the program under evaluation through observation, interview, etc. is considered essential (Charters & Jones, 1973). This is a particularly critical concern when the evaluation involves an "alternative" or "innovative" project compared to a "traditional" program. The last two years of the BESP Level I design focused on comparing BESP students, staff, parents to non-BESP students, staff, parents on a variety of dimensions. Difficulties arose in trying to define BESP and specifying those characteristics which made it unique or alternative.

There are several reasons for this dilemma. First, the programs had undergone changes and refinements throughout their five year existence. Since they were experiments, when certain processes failed they were eliminated; likewise successful features were expanded. This protean quality made the programs difficult to characterize in a consistent manner. Second, with the advent of phase-in and the current financial crisis in the district, the programs are conforming with the mainstream of Berkeley education. BESP programs may be changing but they are moving toward the district status quo. Last year Level I surveyed a sample of the elementary and secondary staff to determine their teaching emphases. Their priorities were found to be similar to those of non-BESP teachers. In observations of instructional methodology (amount of individualization, openness of classrooms, use of team or peer teaching, etc.) it was found that teacher utilization (traditional, team, peer) is independent of ESP/non-ESP categories but that there were more small group instruction, individualization and open teaching practices in the BESP program. However, there was a good deal of variation within the BESP programs in instructional methodology. So by in large BESP schools appear similar to other district schools.

A third difficulty in characterizing the nature of BESP is the wide variety of special programs in Berkeley. BESP is only one of about 40 grants and contracts in the school system; there are a number of federal projects in operation and a number of innovative programs in many schools. Several of the BESP sites

have other programs in operation like Follow-Through and Compensatory Education. The goals and activities of these projects spill over into BESP. The non-BESP sties are also involved in many supplementary programs. Hence, it was impossible to get a "clean" comparison of strictly BESP vs. non-BESP effects, and it is difficult to define BESP in terms distinctly different from non-BESP.

Another problem was encountered in establishing a standard of success for BESP. The success of an educational program is sometimes difficult to determine (what level of accomplishment constitutes success). Criteria for success are simply not well explicated for educational programs. Though the evaluator is admonished to take into account educational or practical significance as well as statistical significance (Scriven, 1974b; Theimer, undated; Cervantes, 1975) specific techniques or models for determining practical significance have yet to be fully detailed.

There are at least five standards that could be used to determine the success of an alternative program:

1. The program accomplishes traditional objectives more effectively than traditional programs (assuming "more effectively" can be established).

2. The program accomplishes new and desirable objectives.

3. Both 1 and 2.

4. The program accomplishes traditional objectives as

effectively as traditional programs.

5. Both 2 and 4.

Any of the five could be used in determining the success of an innovative program, assuming cost factors are also taken into consideration.

For the last years Level I has relied on standard number 4: the program accomplishes traditional objectives as effectively as traditional programs. Given the turbulent history of BESP and the number of changes that have occurred, a program is considered successful if the achievement levels, student, staff and parent satisfaction are comparable to the achievement and satisfaction of those not involved in the special program. Some might argue for more stringent standards for success; number 4 is certainly the most lenient. Given the amount of money spent on BESP and the high hopes, we should perhaps have expected more.

Although BESP Level I has not been successful in developing new evaluation methods or techniques, a few general guidelines can be formulated for assessing alternatives as a result of the BESP experience:

1. The collection of baseline data on program participants is essential. "Traditional" evaluation techniques should <u>not</u> automatically be abandoned, but can be used judiciously to assess the effectiveness of an alternative education program.

2. Enough time should be spent in initial planning so that consistent data can be collected for the duration of a multi-year project. This is not to say that the evaluation component should be locked into a rigid design, but rather certain types of data (e.g., student achievement, participant attitude, etc.) need to be routinely gathered so that program progress can be adequately monitored, trends can be identified, and overall impact at the end of the project can be assessed.

3. Documentation of what the program is actually like (through observations, interviews, surveys, etc.) is crucial in interpreting outcomes and in planning for changes. One technique developed by the district that has proved helpful for multi-goal programs is the use of implementation charts. The charts aid project staff in interrelating project goals, project activities and methods of evaluation. Through a collaborative effort, project staff and evaluator complete a chart focusing on the following: goals of the program for school site; program activities to meet goals, target population goals aimed at; personnel responsible for implementing program activities; method of evaluation; evaluation activities.

4. The distinction between formative and summative evaluation can become muddled in practice. While individual

school sites may want formative feedback, the project director and central administration may require summative data in order to make comparative decisions across programs. The formative-summative distinction, while conceptually useful, is less helpful in practice. The internal evaluator is advised to be prepared for both summative and formative data requests.

5. Finally, evaluating an alternative program poses basically the same problems as evaluating any new or old educational program. The same general processes of establishing evaluator credibility, assisting project staff in formulating measurable goals and objectives, balancing conflicting viewpoints, deciding how and what to measure, etc. are involved in evaluating educational alternatives as are involved in evaluating most educational programs.

## Two Problems

There remain two unresolved problems that have troubled Level I. The first is staff turnover. Changes in staff have been a real difficulty for both the BESP project as a whole and the BESP evaluation. Since its inception in 1971, BESP has had three different Level I staffs, three different BESP directors and numerous changes at the program level. Reasons for staff

turnover may include the frustrations of the job, poor initial selection of staff, or internal politics. In any event, this staff instability has made continuity of evaluation exceedingly difficult. It takes time to establish rapport with new personnel, something not planned for in the original evaluation design. This problem remains unresolved for there are few ways to prevent it. Evaluators need to keep alert to the fact that the project staff, ourselves included, can be very transitory. In setting up scopes of work and evaluation timelines, provisions should be made for the possibilities of staff turnover and time allocated for orientation of new personnel. In addition, perhaps energies and resources could be spent in developing evaluation procedures that are independent of specific personnel so that when (not if) staff changes occur, the disruptions are minimal. This implies thorough documentation of the evaluation procedures and decisions. Unfortunately this area has been largely neglected by BESP Level I.

The second problem involves time. There are two aspects to the time problem. First, timing of the reports to meet the decision maker's needs has been difficult for Level I. It is obvious that data presented after a decision is made is relatively useless. But collecting, analyzing, processing, and summarizing data is a long process. Level I has not been able to respond as quickly as was needed in some instances. The other related aspect of the time problem is the estimation of how long activities will take. This has been an area of weakness in Level I functioning. The length of time to complete an activity has

tended to be underestimated with the result that evaluators overpromised and could not deliver. This, in turn, has affected Level I credibility and resulting requests for evaluation assistance. Although this is, too, a serious problem, it has been more easily remedied than that of staff turnover. Increased familiarity with the evaluation tasks and procedures coupled with experience in completing activities have lessened the severity of this problem. The lesson Level I learned the hard way is to clearly outline in advance what can and can't be provided and to actually allocate time and resources for unanticipated demands and requests.

## Legacy of Level I

Despite its turbulent history and the unresolved problems and the conflicts experienced by Level I, BESP Evaluation has had an impact on the district in several ways. Evaluation personnel are more involved in the early stages of project planning and grant proposal writing than they have been in the past. The role of the evaluator in the district has been expanded from data-gatherer to resource person. Evaluators are involved in critiquing external evaluation reports, writing Request for Proposals for educational audits, summarizing diverse evaluation data in a format suitable for decision making, and consulting with staff on evaluation problems.

In more concrete terms, the impact of Level I can be seen in the rearrangement of the data files and the development of

more meaningful ways of examining achievement test scores. In addition, the district is presently exploring, at the suggestion of BESP Evaluation, the feasibility of developing school profiles which would pull together currently disjointed information so that a school's functioning could be more accurately assessed on a variety of dimensions.

The impact of Level I at the individual schools is less evident. Because of the summative emphasis and lack of formative feedback to the sites, evaluation is still viewed negatively at some sites. The evaluation consciousness of central administration may have been raised, but work still needs to be done at the school sites.

Related to the site personnel's feelings about evaluation, there remain two additional needs: first, more inservice training for the consumers of evaluation so that expectations are realistic and finally, the need for more critical examination of our own evaluation efforts. It is in this spirit that this paper has been written.

## References


Ball, S. & Anderson, S. Professional issues in the evaluation of education/training programs. Princeton, New Jersey: Educational Testing Service, 1975.

Bernstein, I. & Freeman, H. Academic and entrepreneurial research. New York: Russell Sage Foundation, 1975.

Cervantes, R. Issues in assessing non-statistical significance. Paper presented at Experimental Schools Conference, National Institute of Education, Washington, D.C., 1975.

Charters, W. & Jones, J. On the risk of appraising non-events in program evaluation. Educational Researcher, November, 1973, 5-7.

ERIC Abstracts. A collection of ERIC document resumes on assessing educational innovations. ERIC abstract series 23. ED 070 177 1972.

Hein, G. An open education perspective on evaluation. Grand Forks, North Dakota: North Dakota Study Group on Evaluation, 1975.

Jones, C. Berkeley Experimental Schools Project: A narrative history (in progress).

Patton, M. Alternative evaluation research paradigm. Grand Forks, North Dakota: North Dakota Study Group on Evaluation, 1975.

Petrie, H. Do you see what I see? The epistemology of interdisciplinary inquiry. Educational Researcher, February, 1976, 9-15.

Schmuck, R. Alternative schools: contributions of social psychology. Paper presented at American Psychological Association meeting, Chicago, 1975.

Scriven, M. Control of bias. Unpublished manuscript, 1974a.

Scriven, M. Evaluation perspectives and procedures. In W. J. Popham (ed.), Evaluation in Education. Berkeley: McCutchan Publishers, 1974b.

Stufflebeam, D. Criteria for judging evaluation studies. In B. Worthen & J. Sanders (Eds.). Educational evaluation: theory & practice. Belmont, California: Wadsworth Publishing Company, 1973.

Theimer, W. Educational significance of results of evaluation as opposed to statistical significance. Unpublished manuscript, undated.

Worthen, B. Content specialization and educational evaluation: a necessary marriage? Paper presented at American Educational Research Association meeting, Chicago, 1974.